Journal of the Maharaja Sayajirao University of Baroda
ISSN :0025-0422

ANALYSIS OF PLACEMENT PERFORMANCE PREDICTION ON STUDENTS DATA USING MACHINE LEARNING ALGORITHM

B. Kalaiselvi, Research Scholar, Department of Computer Science, Government Arts College, Udumalpet.
Dr. S. Geetha, Assistant Professor, Department of Computer Science, Government Arts College for Women, Puliakulam, Coimbatore.

Abstract:

Machine learning plays key role in the modern higher education methodologies to enhance the performance of the students, predict the outcome of the student and personalized e-learning systems. It helps to produce the proficient graduates and increase the competence of the educators and learners. Academic performance prediction helps to enhance the student's performance in end semester exams and placements. Using classification and clustering techniques from Educational Data Mining (EDM) helps to visualize, classify and predict the academic performance of the students by utilizing the socio-economic and academic data. Prediction of student's performance in placement exams are more complex task for the staffs and placement cell since attaining good marks and getting placement offer from valuable concern will be the final goal for most of the students. Hence better and accurate system is required to predict and enhance the performance in placement exams. This system proposed a decision tree based framework model to predict the placement possibility for the students using passedout student data using j48 classifier. Passed out student data set is used to train the proposed model and also test the trained model to predict the placement possibility. The experimental results shows 87% of prediction accuracy in complete data set and shows the 95% of accuracy in training set after splitting the data set into train (70% of instance) and test (30% of instance).

Keywords:*Placement performance, Decision Tree, j48.*

I. INTRODUCTION

EDM plays significant role to extract the hidden patterns and information from the student's personal, socio-economic and academic data. EDM helps to predict the student's outcome and placement performance, analyses the learning difficulties and requirement of updating the curriculum and teaching strategies. Placements in the period of under graduation completion determine the quality of outcome and standard of the educational institution. Various factors influence the student to obtain good marks in the final exams and obtain the placements.

The proposed decision tree based model is mainly focused on exploring various indicators that have an effect on the placement performance of the students. The extracted hidden information from the training data set (passed out student's details) using J48 classifier is used to generate the decision tree to predict the placement possibility in the test data. It helps to the placement cell, faculties and management to improvise the quality of the students and increase placement opportunities to the students. Extracted hidden information and obtained knowledge from the training data is used for predicting the student's performance in advance.

II. LITERATURE SURVERY

Nuraini A R [1] provides an general idea on the data mining techniques used to predict student performance. In this work reviews on prediction methods such as decision tree, neural network, naïve bayes, k-nearest neighbour and SVM. It uses the Cumulative Grade Point Average and internal assessment of the Malaysian students as a dataset. When compared with all prediction methods neural networks and decision tree is preeminent to predict the students' academic performance.

O Edin [2] considers the student demographic variables, achieved results from high school and from the entrance exam, mind-set towards studying. This research is conducted among first year Students in Department of Economics in university of TUZLA. This paper uses Naïve bayes, multilayer

perceptron and J48 algorithm for classification. Out of these three naïve bayes classifier outperforms than others.

Classification techniques are applied by Delen [3] for imbalanced dataset. Majority of the class have achieved high prediction accuracy while minority of the class having poor performance. In this paper different data balancing techniques are compared to improve the prediction accuracy of poor performance students i.e. over sampling, under sampling and synthetic minority over sampling along with logistic regression, decision trees, neuron networks and support vector machines.

Wulandari [4] categorizes the students using k-means cluster algorithm to expose the hidden pattern and also classifies students based on their demographic data. It uses 300 student's data SPSS16 and forma three clusters such as low performance student, average student and smart student.

Ismail N. H(5) classifies the student academic, demographic and economic data by using decision tree, naive bayes and rule based classification. It shows the Rule based model is best among all techniques by showing highest accuracy value of 71.3%. It is used to predict the success rate of the students in first semester.

III. PROPOSED METHODOLOGY

This work focus on classification and prediction of placement possibilities in the under graduate students data set using data collected from the NGM College, Pollachi with the help of J48 ML classifier. The framework of proposed work is shown in fig. 1.

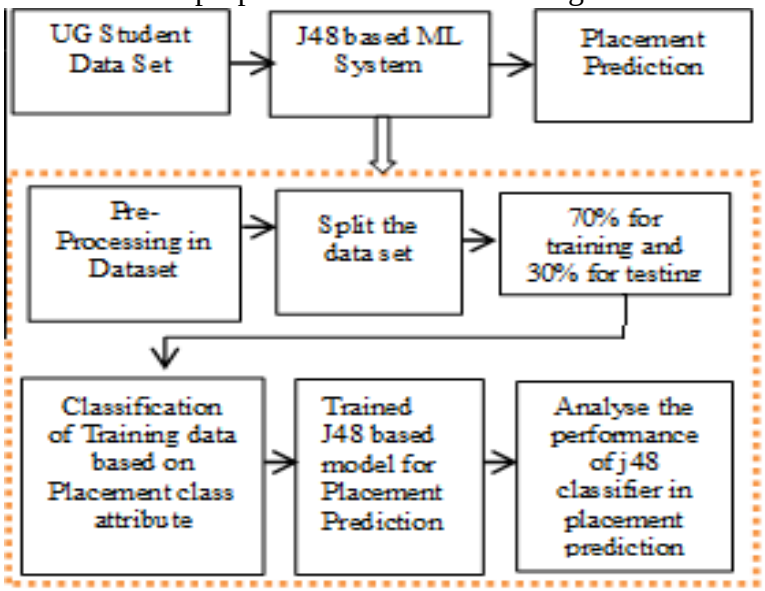

Fig.1. Proposed Framework

Data Set:

The data set consists 124 instances with 8 attributes (Sl.No, Register No, Student name, course, gender, percentage of SSLC, HSC and UG marks and placement status. The data collected from the passed out students from UG IT, CT, CS and B.Com CA departments in the Nallamuthu Gounder Mahalingam College, Pollachi.

Variables	Detail	Possible Values
Sl_No	Serial Number	Numerical Sequence
RegisterNo	Register Number	Alphanumeric Sequence
StName	Student Name	Name of the Student
Course	Course Name	UG Course Name
Gender	Gender	{Male, Female}
SSLCMark	Percentage of Marks in SSLC	{35% to 100%}
HSCMark	Percentage of Marks in HSC	{35% to 100%}

UGMarks	Percentage of UG	{35% to 100%}
Placed	Placed in Campus Interview	{Placed,Not Placed}

Table 1. Attributes used in the data set

Pre-processing:

From the loaded placement data Remove Percentage filter is used to select the 70% instance for training and same filter with invert selection option is used to select remaining 30% instance for test. From the total 124 instances, 87 instances are selected for training and remaining 37 instances for testing.

Classification and Train using J48: ML classifier J48 construct the tree for placement training set in built tree and pruning phases. Category based attributes like course name, percentage of marks and gender are used to build the tree. Information Gain (IG) calculated for the attributes like course, gender, SSLC percentage, H.Sc. percentage, UG Percentage and placement status and assign highest IG value attribute as root note (SSLC Percentage in this work) and form the tree based on splitting the attributes using IG values and checks all out comes fall under the same class or not. If all training instances are placed under the same class, node is represented with single class else determine the splitting attributes to classify and build tree. After tree formation, post / online pruning will be performed in the constructed tree to generate the best tree and solve the outlier and over fitting issues.

Testing and performance measures: After completion of training phase using 48 for selected 87 instances, test data set (partitioned 37 instance data set) is passed to predict the placement possibilities and compare with previous and actual results.Classification accuracy, True positive and false positive rates, Precision and F1 score with Confusion matrix are used to analyse the performance measures of the trained and predicted model.

III IMPLEMENTATION AND RESULTS

The following process are performed in the implementation phase using J48 classifier in Weka;

- Passed out students data classification based on class attribute place (Placed , Not placed),
- Split the passed out students data set into training(70% of instance) and test data set(30% of instances),
- Train the model and test the placement prediction using training and test data sets.

Initially collected 124 passed out students stances are passed to j48 classifier with 10 x folds validation and obtain the classification results and generated tree.

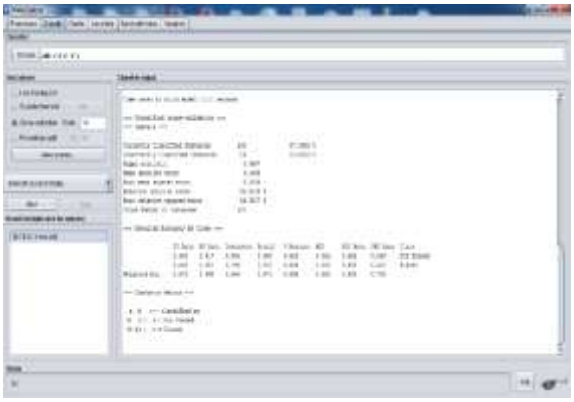

Fig.2 Classification results

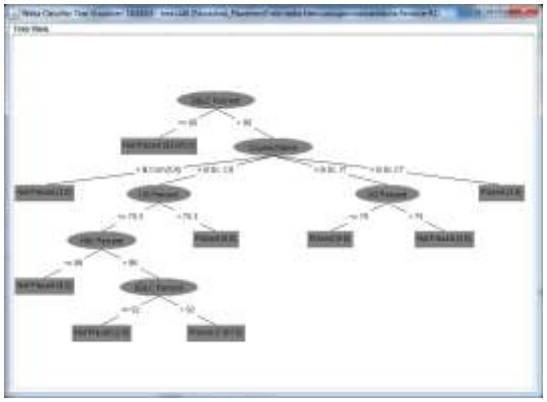

Fig.3 Tree generated by j48

Based on the results (fig.2), from the 124 instances, 108 instances are correctly classified with 87.09% of accuracy. 16 instances are not correctly classified (6 not placed instances are treated as placed and 10 placed instances are treated as not placed). The classifier shows good results in TP rate (0.871), FP Rate(0.348), precision (0.864) and F-measure (0.866).The size of the generated tree

is 15 with 9 leaves. The generated tree is used to classify the placed and not placed students in the created model using training data set. The model takes 0.01 seconds to build.Fig.4. describes margin curve plot, the prediction margin with difference between the probability for the actual class and predicted class. Most of the instances are placed in margin 1, it shows the correct class is predicted by the classifier.

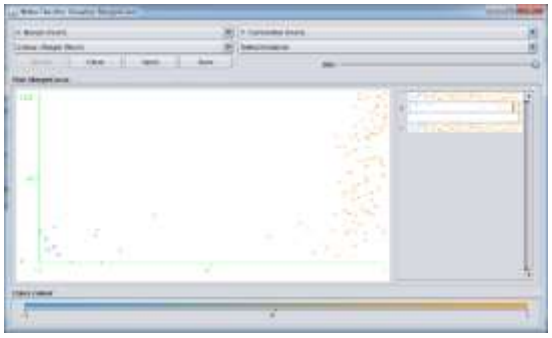

Fig.4 Margin curve plot for complete data set

From the total 124 instances, 87 instances are selected for training and remaining 37 instances for testing using split percentage and invert selection method. The selected 87 instance data set is used to train the model. The following fig.5 shows split placement data set into train and test random percentage and invert selection.

Fig.5. Data Set split

Fig.6. Save Predicted result

The above fig.6 depicts store the test data with predicted results

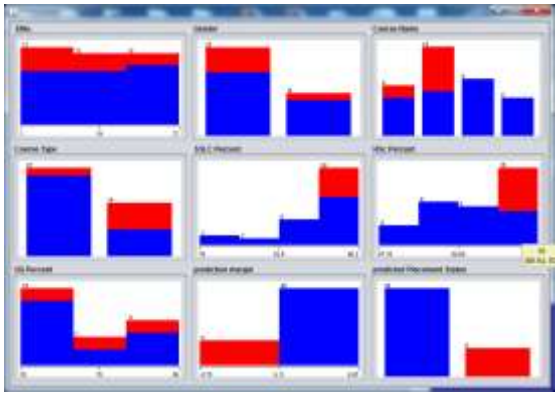

Fig.7. Visualized representation of predicted result data set

IV DISCUSSION

Out of 124 students data, the system predicted 108 instances correctly in the classification process carried out by j48 classifier based on placement status (placed, not placed). J48 classifier produced 87.09% accuracy with 0.5587 as Kappa statistic value, 0.1629 as Mean absolute error and 0.3436 as root mean square error. The j48 classifier builds the placement prediction model in 0.01 seconds only. It shows good accuracy with least running time.

After splitting the test data set from original placement data set, the test data set consists of 13 placed students and 24 non placed students. This model predicts 28 students are not placed and 9 students are placed. It reveals that some variations found in the placement prediction process, remaining nearly 13% of prediction accuracy should be removed from the prediction system and make it as 100% prediction accuracy.

CONCLUSION AND FUTURE SCOPE

This prediction model closely met the objective of this prediction model with 87% of placement prediction accuracy in the passed out student's placement training and test data set. This system yields good accuracy rate and it helps to the placement cell and also staffs to improve the students skill set to make them as placement offer received students. . To enhance the accuracy of j48 classifier, entropy value is modified using Rank correlation algorithm and attributes are selected using CFS Subset algorithm.

ACKNOWLEDGEMENT

I convey my sincere thanks to the management of Nallamuthu Gounder Mahalingam College, Pollachi for allowing me to carry out this research work by funding with SEED MONEY.

REFERENCES

1. Amirah Mohammed S., Wahidah H, & Nuraini A R, (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/J.PROCS.2015.12.157
2. O Edin. ;,& S Mishra. (2012). Data Mining Approach for Predicting Student Performance. In *Economic Review: Journal of Economics and Business* (Vol. 10, Issue 1). http://hdl.handle.net/10419/193806
3. Thammasiri, D., Delen, D., Meesad, P., &Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, *41*(2), 321–330. https://doi.org/10.1016/j.eswa.2013.07.046
4. Harwati, Alfiani, A. P., &Wulandari, F. A. (2015). Mapping Student's Performance Based on Data Mining Approach (A Case Study). *Agriculture and Agricultural Science Procedia*, *3*, 173–177. https://doi.org/10.1016/j.aaspro.2015.01.034
5. Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, *9*(129), 6415–6426. https://doi.org/10.12988/ams.2015.53289